

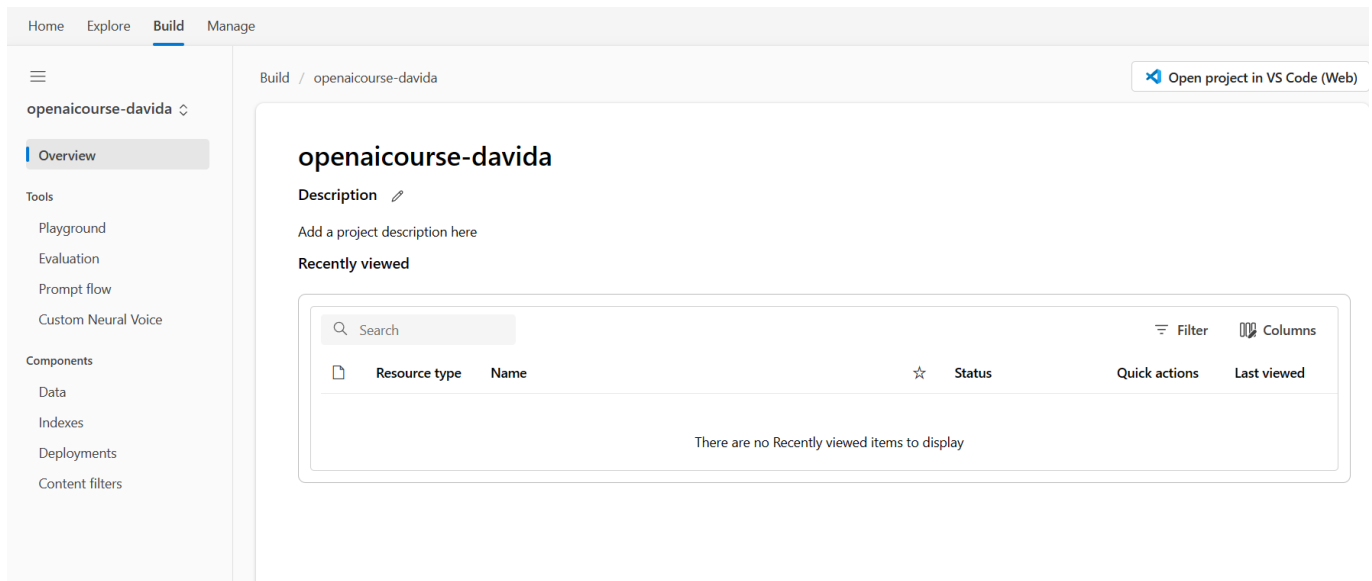
# Navigate Azure AI Project Options

This guide helps you navigate through the different options within your Azure AI Project.

**Duration: 10 to 15 minutes**

## Project Overview

Start by returning to the build section and selecting your project.



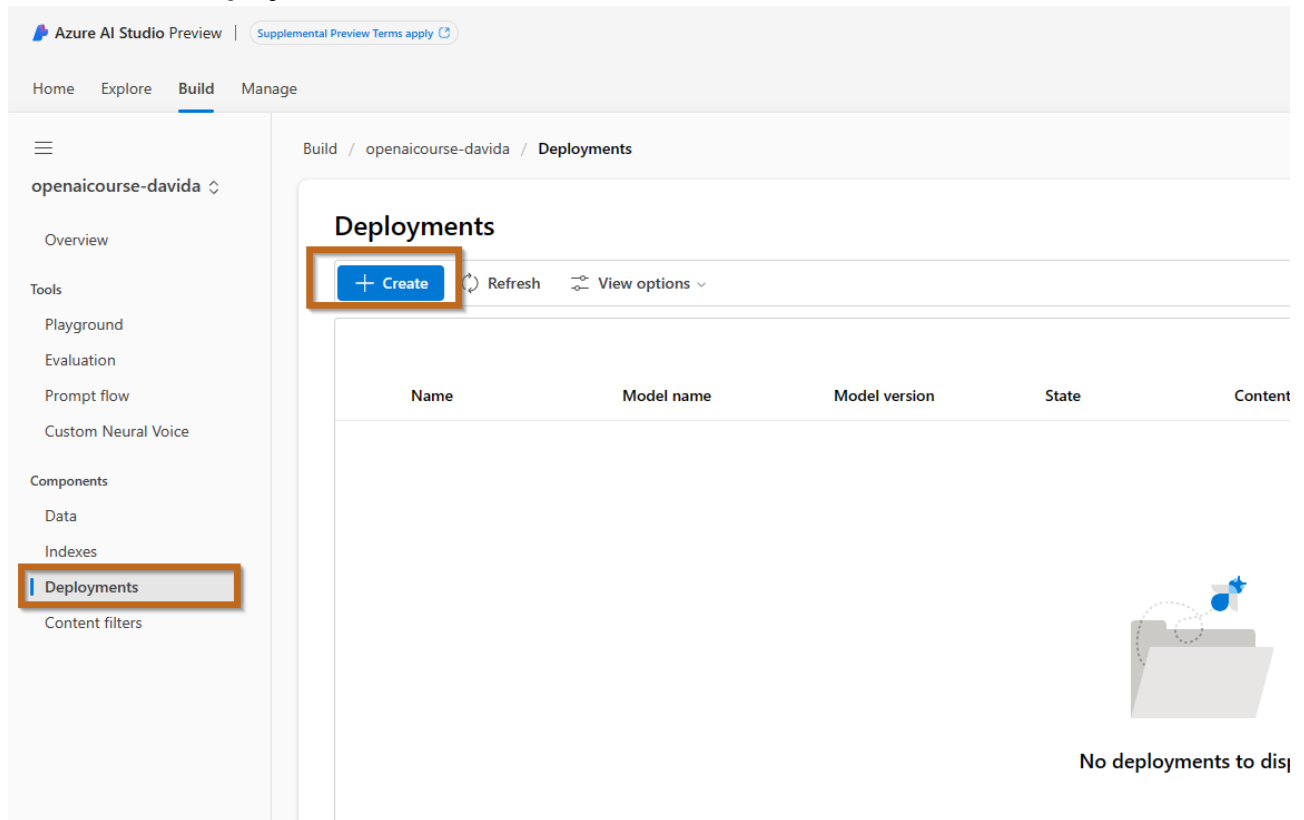
## Key Sections in Your Project

1. **Overview:** A general overview of your project's status and settings.
2. **Tools:** Tools available for your project:
  - **Playground:** Test your model and chat session here.
  - **Evaluation:** Evaluate model performance with standard metrics to choose the best version for your needs.
  - **Prompt Flow:** A development tool for streamlining AI application development with Large Language Models (LLMs).
  - **Custom Neural Voice:** Design a unique voice persona and manage voice talents, datasets, models, tests, and endpoint connections.
3. **Components:** Essential components for your project:
  - **Data:** Import data for your models.
  - **Indexes:** Create indexes to customise generative AI responses.
  - **Deployments:** Set up the deployment infrastructure for external model consumption.
  - **Content Filters:** Manage content filtering configurations that work alongside core models.

## Deployments

Before utilizing the playground, you need to create a deployment.

## 1. Create a New Deployment:



Azure AI Studio Preview | Supplemental Preview Terms apply

Home Explore Build Manage

Build / openaia-course-davida / Deployments

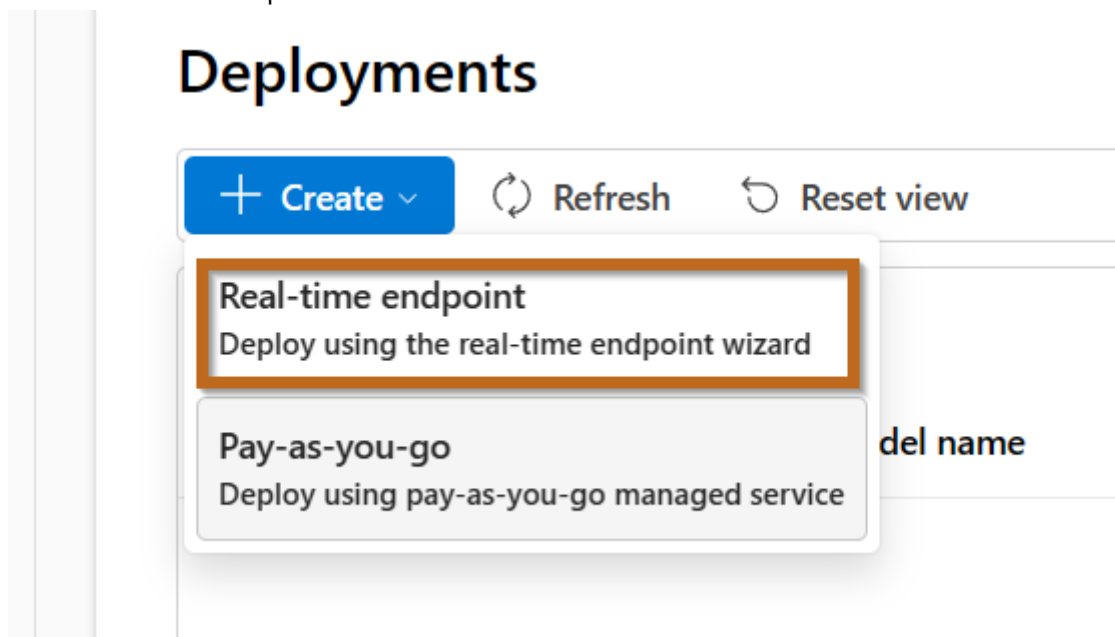
### Deployments

**+ Create** Refresh View options

Name	Model name	Model version	State	Content
------	------------	---------------	-------	---------

No deployments to display

- Select Real-Time Endpoint



## Deployments

**+ Create** Refresh Reset view

**Real-time endpoint**  
Deploy using the real-time endpoint wizard

**Pay-as-you-go**  
Deploy using pay-as-you-go managed service

- Select the model you wish to deploy, any gpt model will work correctly for the course.

## Select a model

Choose a model to create a new deployment. For flows and other resources, create a deployment from their respective list. [Go to model catalog.](#)

Models: 1616

Collections

Tasks



Show description

Search



gpt-35-turbo-16k

Chat completions



gpt-35-turbo-instruct

Chat completions



gpt-4

Chat completions



babbage-002

Completions



davinci-002

Completions



gpt-35-turbo

Chat completions



gpt-4-32k

Chat completions



text-embedding-ada-002



Select a model to see description

Confirm

Cancel

- Confirm and deploy the model.

## Deploy model



Unable to get Azure OpenAI quota information. You can still try deploying, but your deployment may fail if you do not have enough quota.

Deployment name \*



gpt-4-32k

Select model \*

gpt-4-32k



Advanced options



Deploy

Cancel

- Leave the options in the following window by default and click Deploy.

The screenshot shows a 'Deploy model' window with the following fields and options:

- Deployment name \***: A text input field containing 'gpt-4'.
- Select model \***: A dropdown menu showing 'gpt-4'.
- Model version**: A dropdown menu showing '0125-Preview'.
- Advanced options**: A section header with an upward arrow icon.
- Content filter ⓘ**: A dropdown menu showing 'Default'.
- Deployment type**: A dropdown menu showing 'Standard'.
- 80K tokens per minute quota available for your deployment**: A grey informational banner.
- Tokens per Minute Rate Limit (thousands) ⓘ**: A slider control with a value of 10K.
- Corresponding requests per minute (RPM) = 60**: Text displayed below the slider.

## 2. Repeat for Additional Models:

- Select another model.

### Select a model

Select a model that you want to deploy. This can be consumed as a base model or used inside a flow.

Name	Collection	Task
<input checked="" type="checkbox"/> gpt-35-turbo	Azure OpenAI	Chat completion
<input type="checkbox"/> gpt-4	Azure OpenAI	Chat completion
<input type="checkbox"/> gpt-35-turbo-16k	Azure OpenAI	Chat completion
<input type="checkbox"/> gpt-4-32k	Azure OpenAI	Chat completion
<input type="checkbox"/> text-embedding-ada-002	Azure OpenAI	Embeddings
<input type="checkbox"/> mistralai-Mistral-7B-v01	Curated by Azure AI	Text generation
<input type="checkbox"/> mistralai-Mistral-7B-Instruct-v...	Curated by Azure AI	Chat completion

< Prev Next >

Confirm Cancel

- Select the model you wish to deploy, any gpt model will work correctly for the course.

### Select a model

Choose a model to create a new deployment. For flows and other resources, create a deployment from their respective list. [Go to model catalog.](#)

Models: 1616 Collections Tasks Show description

☒ gpt-35-turbo-16k  
Chat completions

☒ gpt-35-turbo-instruct  
Chat completions

☒ gpt-4  
Chat completions

☐ babbage-002  
Completions

☐ davinci-002  
Completions

☒ gpt-35-turbo  
Chat completions

☒ gpt-4-32k  
Chat completions

☐ text-embedding-ada-002

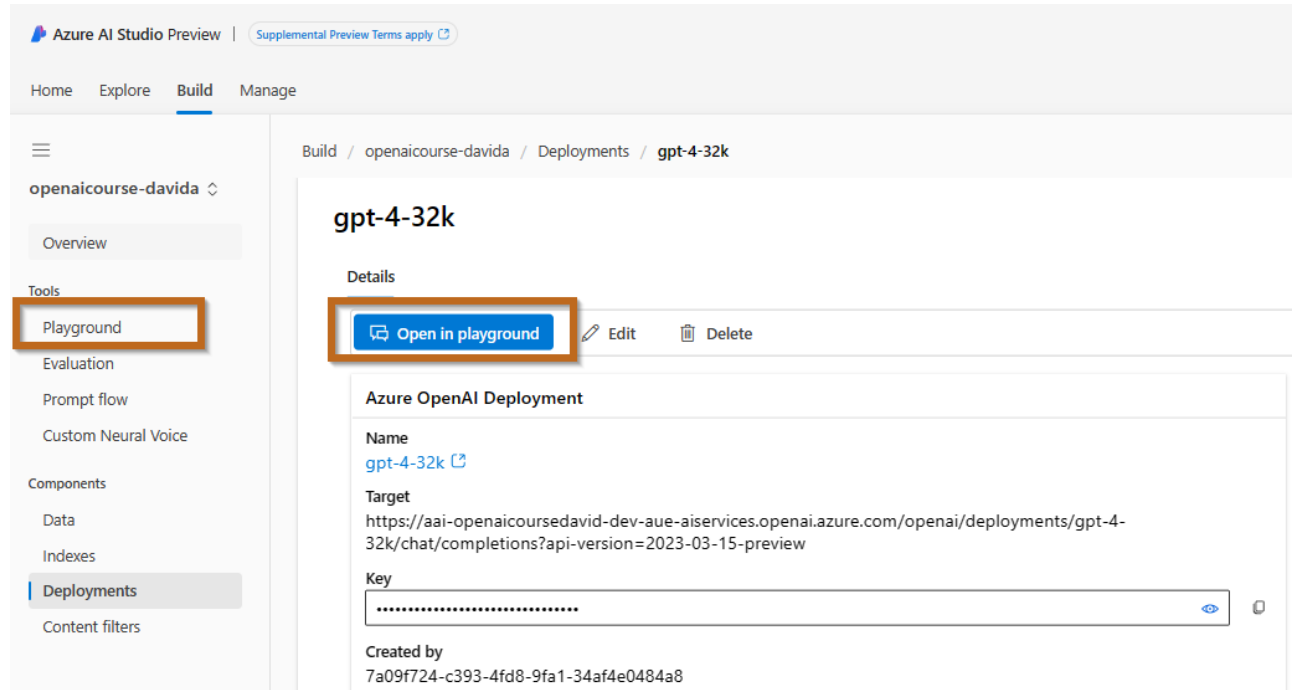
Select a model to see description

Confirm Cancel

Confidential - Actionable Data Analytics Pty Ltd (ABN 79 649 102 536) - 5 / 7

**Note:** Creating multiple models can help manage heavy traffic and provide alternatives.

### 3. Navigate to the Playground:



**Tip:** Deploying multiple models is useful for testing solutions and choosing the right model.

## Using the Playground

The *Chat* playground offers a chatbot interface for your deployment models, utilising the *ChatCompletions* API.

1. **Explore Options:** Familiarise yourself with various features in this section. We will delve into more details in the next exercise.

**Info:** The top right corner of the playground displays the API in use.

## Grounding

As discussed previously, grounding ensures responses are based on real-world knowledge from extensive training data, but limitations exist due to data cutoff dates and potential inaccuracies from training patterns.

Using the following prompt:

What is your cutoff date? what's the latest information you have been trained on?

Build / std-000\_ai / std-000-9774 / Playground

Open project in VS Code (Web) ▾

## Playground

Manual evaluation Deploy to a web app Import Export View Code Prompt samples ...

Mode Chat ▾

## System message

Add your data

Examples

Apply changes

Reset to default

## System message ⓘ

You are an AI assistant that helps people find information.

Variables

{x} Add variable

Replay chat

Clear chat

Playground Settings

☐ Show JSON

what is your cutoff date? what's the latest information you have been trained on?

My training includes information up to April 2023. This means the latest updates, news, or developments occurring after this date might not be included in my responses. If you have any questions or need information up to that point, feel free to ask!

Type user query here. (Shift + Enter for new line)

Write the questions here.

## Deployment

Parameters

## Deployment \*

gpt-4 ▾

## Session settings

Past messages included ⓘ

 10